統計数理研究所彙報
第32巻 第2号 (1984)

多変量統計量の正規化・分散安定化変換

統計数理研究所 小 西 貞 則

(1984年8月 受付)

1. は じ め に

統計量を何らかの関数で変換し，この変換された統計量に基づいて各種統計的推測を行う．この考え方の古典的かつ実際的な例として，Fisher (1921)によって提唱された相関係数に対する逆双曲線関数による変換，すなわち一般にFisherの z 変換と呼ばれているものが上げられる．いま母集団相関係数 ρ をもつ二変量正規母集団からの大きさ $N=n+1$ 個の無作為標本にもとづく標本相関係数を r とする．Fisherの z 変換は，

$$z(r)=\frac{1}{2}\log\{(1+r)/(1-r)\}$$

で与えられ，$z(r)$ は n が十分大きいとき，平均 $z(\rho)$，分散 $1/(n-2)$ の正規分布で近似することができる．

実用上この z 変換が用いられるのは，次の二つの異なる理由からである．すなわち変換を行うことによって，

(i) 漸近分散がパラメータ ρ に依存しない形にできる．

(ii) 正規分布にもとづく近似式の精度を改善することができる．

一般に(i)の変換は分散安定化変換 (Variance stabilizing transformation)，(ii)の変換は正規化変換 (Normalizing transformation) とよばれている．

Hotelling (1953) は，標本相関係数の分布およびFisherの z 変換について，理論的研究を行い示唆に富んだ多くの結果を与えたが，この論文に対する討論の中で，Kendall, Anscombe 等の人達が一様に指摘していることは，「なぜ z 変換によって，(i), (ii)の性質が同時に満たされるのか理論的に明白でない.」という点である．本稿の第一の目的は，この変換の意味を理論的に明らかにすることにある．

一般に統計量の分散安定化変換は(i)によって明確に定義され，その変換形を求める方法は，Bartlett (1948)によって形式的ではあるが，分散分析の問題に関連して紹介された．これに対して，正規化変換は(ii)によって必ずしも明確に定義されているとはいえず，正規化変換を求める方法も統一的には議論されていなかった．この問題に対して，Konishi (1981) は正規化変換を〝極限分布である正規分布への収束速度の改善〟という観点から把え，一つの統一的な意味づけを与えた．

本稿の第二の目的は，Konishi (1981) に従って一般に漸近正規性を有する統計量の正規化変換を，近似誤差を評価する観点から把えその定義を与え，種々の統計量の正規化変換を導出することにある．さらに，分散安定化変換と正規化変換の関係についても考察する．

なお，最近 Efron (1982)が，本稿とは別の観点から，同じテーマを扱っているので，参考になると思う．

2. 正規化・分散安定化変換

2.1. Fisher の z 変換 (Fisher (1921))

母集団相関係数 ρ をもつ二変量正規母集団からの大きさ $N=n+1$ 個の無作為標本に基づく標本相関係数を r とする．$n\to+\infty$ のとき

(2.1) $$\sqrt{n}(r-\rho)/(1-\rho^2) \qquad (\mathrm{R})$$

の極限分布は平均 0，分散 1 の正規分布 $N(0, 1)$ であり，したがってこの場合 r の漸近分散は $(1-\rho^2)^2/n$ となり母集団相関係数 ρ に依存している．いま $f(r)$ を標本相関係数 r の 1 対 1 連続微分可能な関数とする．このとき

(2.2) $$\sqrt{n}\,\{f(r)-f(\rho)\}/\{(1-\rho^2)f'(\rho)\}$$

の分布は，$n\to+\infty$ のとき $N(0, 1)$ に収束することが証明される（例えば Anderson (1958, p. 77)）．したがって変換を行うことによって漸近分散がパラメータ ρ に依存しない形にするには，(2.2) から直ちに微分方程式

(2.3) $$(1-\rho^2)\frac{df}{d\rho}=1$$

の解を求め，この関数に基づく変換を行えばよいことに気付く．微分方程式 (2.3) の解は $f(r)=\frac{1}{2}\log\{(1+r)/(1-r)\}(=z(r))$ で，したがって z 変換が標本相関係数 r に対する分散安定化変換であることがわかる．

標本相関係数に対する z 変換は同時に正規化変換と呼ばれ，(2.1) に基づく近似精度を著しく改善するということからも用いられている．実用上しばしば用いられるのが，

(2.4) $$\sqrt{n-2}\{z(r)-z(\rho)\} \qquad (\mathrm{Z}_1)$$

の分布を $N(0, 1)$ で近似する方法である．これは $z(r)$ の漸近分散が

$$\sigma_z^2=n^{-1}\left\{1+n^{-1}\left(2-\frac{1}{2}\rho^2\right)+\cdots\cdots\right\}$$
$$\approx n^{-1}(1-2n^{-1})^{-1}=(n-2)^{-1}$$

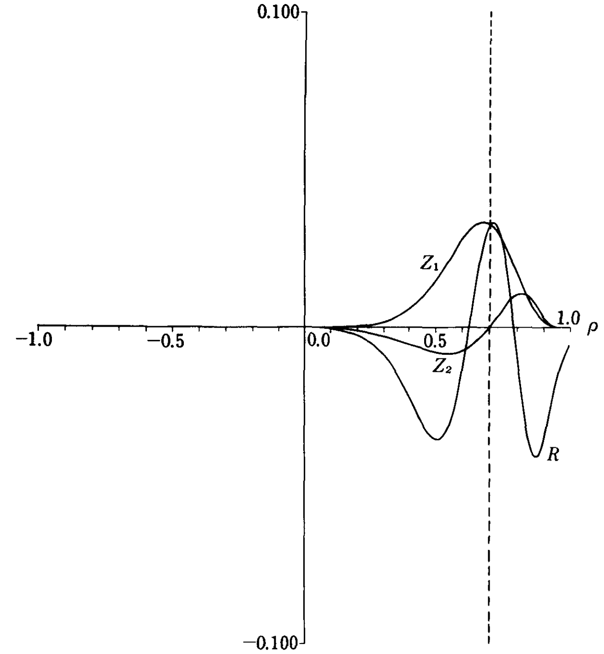
と近似されることによるもので $|\rho|$ が 0 に近ければ多少の修正にはなっている．これに対して $z(r)$ の漸近平均をオーダー n^{-1} の項まで取り入れると

(2.5) $$\sqrt{n}\left\{z(r)-z(\rho)-\frac{\rho}{2n}\right\} \qquad (\mathrm{Z}_2)$$

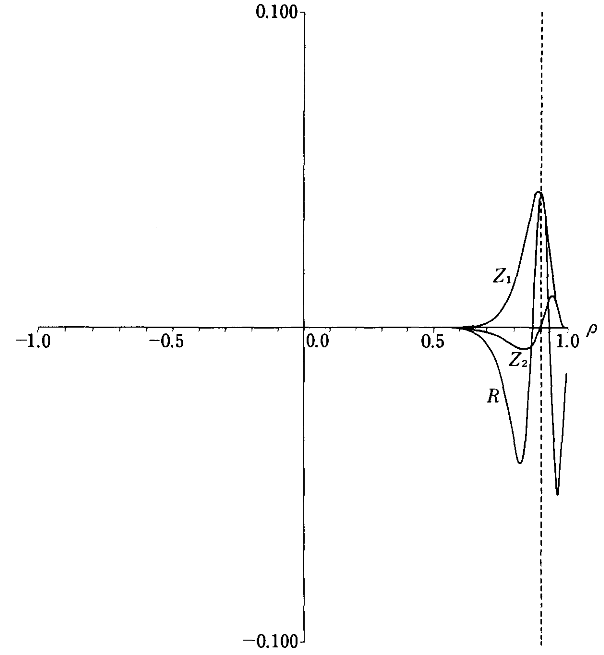
を $N(0,1)$ で近似することができる．そこで，(2.1), (2.4), (2.5) で与えた三つの近似式を各々 R, Z_1, Z_2 とし，変換を行うことによって，どの程度近似精度が改善されるものなのかを比較してみることにする．

図 1 は確率分布の値 $\Pr(r<r_0)$ を三通りの近似式を用いて計算し，それらと真の値との差を表したものである．これを見ると標本相関係数 r の極限分布に基づく近似式 R の精度は近似式 Z_2 によって r の定義域全体に渡ってかなり改良されているといえるが，実用上しばしば用いられる近似式 Z_1 は，分布の両裾において変換の効果は現れているが，母集団相関係数 ρ の近傍ではその効果は現れているとは言えない．そこで，z 変換を施すことによってなぜ近似精度が改善されるのか，また近似式 (2.4) と (2.5) の近似精度の違いは，どこに起因するのかを理論的に考察すると共に，第一章で述べた (i), (ii) の性質を同時に満たす変換が z 変換であることを自然な形で導出する．

いま $f(r)$ を標本相関係数 r の 1 対 1 かつ 2 回連続微分可能な関数とする．$n\to+\infty$ とする



(a) 標本数=20
母集団相関係数=0.7



(b) 標本数=20
母集団相関係数=0.9

図1.　確率分布の値 Pr ($r<r_0$) を (2.1) 式 (R), (2.4) 式 (Z_1), (2.5) 式 (Z_2) を用いて近似したときの誤差曲線：誤差＝（近似値ー真の値）

と $f(r)$ を規準化した (2.2) の分布が標準正規分布関数 $\Phi(x)$ に収束するということは確率分布の値を

$$\Pr[\sqrt{n}\{f(r)-f(\rho)\}/\{(1-\rho^2)f'(\rho)\}<x]\approx\Phi(x)$$

で近似できることを表す.

これは, $f(r)$ に対する分布の漸近展開において, $n\to+\infty$ にすることで, 極限分布 $\Phi(x)$ に続く項を無視して近似していることを示す. そこで有限な n に対して, 標準正規分布関数 $\Phi(x)$ に続く第二項目を求めると

$$(2.6)\quad \begin{aligned}&\Pr[\sqrt{n}\{f(r)-f(\rho)\}/\{(1-\rho^2)f'(\rho)\}<x]\\&=\Phi(x)-\frac{1}{\sqrt{n}}\left[\frac{1}{2}\rho-\left\{\rho-\frac{1}{2}(1-\rho^2)f''(\rho)f'(\rho)^{-1}\right\}x^2\right]\phi(x)+O(n^{-1})\end{aligned}$$

となる. ここに $\phi(x)$ は標準正規密度関数とする.

いま標本相関係数の確率分布の値 $\Pr(r<r_0)(|r_0|<1)$ を (2.6) 式を用いて近似するとき, 一般に狭義単調関数 $g(r)$ に対しては $\Pr(r<r_0)=\Pr\{g(r)<g(r_0)\}$ であるから, (2.6) 式に含まれる変数 x に

$$x=\sqrt{n}\{f(r_0)-f(\rho)\}/\{(1-\rho^2)f'(\rho)\}$$

という値を与える. したがって x の値は $r_0=\rho$ の時 0 で ρ から離れ分布の両裾に行くに従って大きな値をとることになる. いま確率分布の値を (2.6) 式の第一項 $\Phi(x)$ を用いて正規近似したとすると, その近似誤差は $\Phi(x)$ に続く第二項目の $1/\sqrt{n}$ の項の大きさに主として依存するものと考えられる. $1/\sqrt{n}$ の項を見ると $x=0$ のとき $\rho/2$ が残り従って ρ の近傍ではどのような変換を行っても極限分布 $\Phi(x)$ で近似する限り, それほど近似精度を向上させることにはならないことがわかる. これが分布の裾へ行くにつれて x は大きくなり $1/\sqrt{n}$ の項に含まれる $\{\rho-(1-\rho^2)f''(\rho)f'(\rho)^{-1}/2\}x^2$ の項も無視できなくなる. そこで任意の x に対してこの項を消去するため x^2 の係数を 0 とする変換, すなわち微分方程式

$$(2.7)\quad \rho-\frac{1}{2}(1-\rho^2)f''(\rho)f'(\rho)^{-1}=0$$

を満たす変換式を考えることにする. この微分方程式の解は $z(r)=(1/2)\log\{(1+r)/(1-r)\}$, すなわち Fisher の z 変換で, 微分方程式 (2.3) の解として与えられる分散安定化変換と一致していることがわかる. いま (2.6) 式において $f(r)=z(r)$ とおいて, z 変換の分布の漸近展開を $1/\sqrt{n}$ の項まで求めると

$$\Pr[\sqrt{n}\{z(r)-z(\rho)\}<x]=\Phi(x)-\frac{1}{\sqrt{n}}\left(\frac{1}{2}\rho\right)\phi(x)+O(n^{-1})$$

となる. そこで $1/\sqrt{n}$ の項に残った $\rho/2$ を消却することによって極限分布 $\Phi(x)$ に続く第二項のオーダーを $O(1/\sqrt{n})$ から $O(n^{-1})$ へと改善することを試みる. そのため $z(r)$ を, 修正項 c を用いて $z(\rho)+c/n$ で規準化すると

$$\Pr\left[\sqrt{n}\left\{z(r)-z(\rho)-\frac{c}{n}\right\}<x\right]=\Phi(x)-\frac{1}{\sqrt{n}}\left(\frac{1}{2}\rho-c\right)\phi(x)+O(n^{-1})$$

がえられる. そこで $c=\rho/2$ とおくと

$$(2.8)\quad \Pr\left[\sqrt{n}\left\{z(r)-z(\rho)-\frac{\rho}{2n}\right\}<x\right]=\Phi(x)+O(n^{-1})$$

となり, $1/\sqrt{n}$ の項を消却することができた. これに対して (2.1), (2.4) で与えられる近似式の極限分布への収束のオーダーは各々

$$(2.9)\quad \Pr\{\sqrt{n}(r-\rho)/(1-\rho^2)<x\}=\Phi(x)+O(1/\sqrt{n}),$$

$$(2.10)\quad \Pr[\sqrt{n-2}\{z(r)-z(\rho)\}<x]=\Phi(x)+O(1/\sqrt{n})$$

であることに注意する. (2.9) 式と (2.10) 式の違いは, (2.9) 式が $1/\sqrt{n}$ の項に $-\rho x^2$ の形で x^2

の項を含むのに対して，(2.10) 式は，$1/\sqrt{n}$ の項には x^2 の項は含まれず，これが図 1 の近似精度の違いになって現れてきたわけである．

以上の考察から，正規化変換が存在するとは，変換によって (2.8) 式のタイプの近似が可能なとき，すなわち "統計量に何らかの変換を施すことによって極限分布への収束速度を改善することができるとき" と定義することにする．

分散安定化変換および正規化変換が同じ変換によって与えられるには，(2.3) と (2.7) の二つの微分方程式を同時に満たす解が必要となり，標本相関係数の場合

$$\rho-\frac{1}{2}(1-\rho^2)\left\{\log\frac{df}{d\rho}\right\}'=\rho-\frac{1}{2}(1-\rho^2)\{\log(1-\rho^2)^{-1}\}'=0$$

が成立している故，同じ解 $z(r)$ が得られたわけである．

標本相関係数の分布については，より簡単でより高い近似精度を有する式の導出を目的として，多くの人達によって研究されてきた（例えば，Ruben (1966), Kraemer (1973) など）．Konishi (1978) は，標本相関係数の分布の漸近展開のオーダー $1/n$ の項を補正し，過去に提唱されたものと比べて簡単でより近似精度の高い式を導出した．

また最近 Konishi (1984) は k 変量正規母集団からの無作為標本に基づく級内相関係数の最尤推定量に対して用いられている z 変換（Fisher (1958, Chap. 7)）は，任意の次元 k に対して分散安定化変換であるが，同時に正規化変換となるのは $k=2$ のときだけであることを示し，$k\geqq 3$ のときの正規化変換をベキ変換の形で導出した．

2.2 一般化正規化変換

標本相関係数に対する z 変換のような変換が存在すれば，極限分布への収束速度が改善され少ない標本数でも比較的良い近似値を得ることができる．そこで，一般に漸近正規性を有する統計量に対して正規化変換を導出するための一般論を与える．

確率変数 T_n の分布は，パラメータ $n, \theta=(\theta_1, \cdots, \theta_p)'$ に依存するものとし，また $\mu(\theta), \sigma(\theta)$ を

$$\sqrt{n}\{T_n-\mu(\theta)\}/\sigma(\theta)$$

の分布が $n\to+\infty$ のとき標準正規分布関数に収束するように見い出すことができたとする．いま $f(T_n)$ を $T_n=\mu(\theta)(=\mu)$ の近傍で 1 対 1 かつ 2 回連続微分可能な関数とする．このとき $f(T_n)$ の分布の $1/\sqrt{n}$ の項までの漸近展開が

$$\begin{aligned}&\Pr[\sqrt{n}\{f(T_n)-f(\mu)-c/n\}/\{\sigma(\theta)f'(\mu)\}<x]\\ &=\Phi(x)-\frac{1}{\sqrt{n}}[\{a_1(\theta, f'(\mu), f''(\mu))-c/\sigma(\theta)f'(\mu)\}\\ &\quad+a_2(\theta, f'(\mu), f''(\mu))x^2]\phi(x)+O(n^{-1}),\end{aligned}\tag{2.11}$$

の形で得られたとする（補遺を参照）．c はパラメータに依存する関数 $f(T_n)$ の漸近平均に対する補正項である．多変量解析における漸近正規性を有する統計量に対しては，その多くが (2.11) の形で漸近展開が求まることに注意する．

分散安定化変換は (2.11) より直ちに微分方程式

$$\sigma(\theta)f'(\mu(\theta))=1$$

の解として与えられることがわかる．もちろんこの変換は必ずしも良い近似を与える変換へとは結びつかない．

漸近展開式 (2.11) における $1/\sqrt{n}$ の項を任意の x に対して 0 とするためには，まず x^2 の係数を 0 とする関数を求める必要がある．そこで微分方程式

$$a_2(\theta, f'(\mu), f''(\mu))=0$$

を解くことによって得られる解 f_0 を用いて，統計量 T_n を変換する．すなわちここで変換形 $f_0(T_n)$ が定まる．次に

$$c_0=\sigma(\theta)f_0'(\mu)a_1(\theta,\ f_0'(\mu),\ f_0''(\mu))$$

と置くと

$$\begin{aligned}&\Pr[\sqrt{n}\{f_0(T_n)-f_0(\mu)-c_0/n\}/\{\sigma(\theta)f_0'(\mu)\}<x]\\&=\Phi(x)+O(n^{-1})\end{aligned}\tag{2.12}$$

となり，$1/\sqrt{n}$ の項は消え収束のオーダーを改善することができた．このように，(2.12)が成り立ち，かつ $a_2=0$ の解 f_0 があるとき，確率変数 T_n の正規化変換は存在し，変換形は $f_0(T_n)$ であると定義する．

標本相関係数に対する z 変換のように，分散安定化変換が同時に正規化変換であるためには，微分方程式 $\sigma(\theta)f'(\mu)=1$ の解 f_0 に対して $a_2(\theta,\ f_0'(\mu),\ f_0''(\mu))=0$ である場合に限ることが以上のことからわかる．

補遺で与えた漸近展開(2.11)の導出法からわかるように $1/\sqrt{n}$ の項に含まれる x^2 の係数 $a_2(\theta,\ f'(\mu),\ f''(\mu))$ は，$\sqrt{n}\{f(T_n)-f(\mu)\}/\{\sigma(\theta)f'(\mu)\}$ の三次のキュムラントの第一項を定数倍したものである．これを0にするということは，変換を行うことによって分布の歪みを補正し，より対称な分布に近づける操作を行っていることを示す．しかし近似精度の改善という点から見たときこの変換の効果は，分布の両裾においては現れるが定義域全体にわたるものではない．(2.12)の形で正規化変換が得られたとき，初めて極限分布である標準正規分布への収束速度がかわり，近似式の精度が改善されたといえる．

最近，「数式処理システム」すなわち「数式を記号のまま用いて，計算機で処理することが可能なシステム」を活用し，統計量の分布の高次漸近展開を導出する研究が行われつつある(Niki and Konishi (1984 a))．この際，統計量を上述の方法で変換し，変換された統計量の分布の漸近展開を求めることによって，収束性が著しく改善されるということが，Niki and Konishi (1984 b) の研究で示された．

3.　多変量統計量の正規化変換

2.2節で与えた，一般に漸近正規性を有する統計量の正規化変換を求める方法を用いて，多変量解析における各種統計量の正規化変換を導出し，更に分散安定化変換との関係について検討する．尚，3.2節(i)の標本分散共分散行列の固有根，3.3節の正準相関係数および3.4節(i)の重相関係数の正規化変換は Konishi (1981) にもとづくものである．

3.1.　二次形式統計量

確率ベクトル $y=(y_1, y_2, \cdots, y_p)'$ は平均ベクトル μ，分散共分散行列 Σ をもつ p 変量正規分布 $N_p(\mu, \Sigma)$ に従うとする．二次形式統計量は，対称行列 A に対して $y'Ay$ で定義される．正規性の仮定のもとで導出された各種統計量の多くがこの二次形式統計量の形に帰着されることから，その分布は精密分布，近似分布の両面に渡って幅広く研究されてきた(例えば，Johnson and Kotz (1970, Chap. 29) を参照)．

二次形式統計量 $y'Ay$ の分布は，適当な変数変換を行うことによって

$$Q_p(x)=\sum_{j=1}^{p}\lambda_j(x_j+a_j)^2\tag{3.1}$$

の分布を求める問題に帰着される．ここに $x=(x_1,\ x_2,\ \cdots,\ x_p)'$ は $N_p(0,\ 1)$ に従う p 変量確率ベクトル，$\lambda_j,\ a_j(j=1, 2, \cdots, p)$ は有限の定数でここでは，$\lambda_j>0$ であると仮定する．$Q_p(x)$ の

r 次のキュムラント χ_r は

$$\chi_r=2^{r-1}(r-1)!\,\theta_r$$

で与えられる．ただし

$$\theta_r=\sum_{j=1}^{p}\lambda_j^r(1+ra_j^2) \tag{3.2}$$

とおく．このとき $p\to+\infty$ により $Q_p(x)=Q_p$ の平均 $\theta_1\to+\infty$ とすると，$\theta_r/\theta_1=O(1)$ であることに注意すれば，

$$\frac{Q_p-\theta_1}{\sqrt{2\theta_2}}=\sqrt{\theta_1}\,\frac{(Q_p/\theta_1)-1}{\sqrt{2(\theta_2/\theta_1)}}$$

の分布は $N(0,1)$ に収束することが示される．すなわち $\theta_1\to+\infty$ とするとき，統計量 Q_p/θ_1 が漸近正規性をもつことが示される．そこで Q_p/θ_1 の正規化変換を 2.2 節に従って求めることにする．

いま $\omega_r=\theta_r/\theta_1$ とおくと，(2.11) 式に対する Q_p/θ_1 の関数の分布の漸近展開は，

$$\begin{aligned}&\Pr\left[\sqrt{\theta_1}\left\{f(\alpha Q_p/\theta_1)-f(\alpha)-\frac{1}{\theta_1}c\right\}/\tau<x\right]\\&=\Phi(x)-\frac{1}{\sqrt{\theta_1}}\left[-\frac{\sqrt{2}}{3}\omega_3\omega_2^{-3/2}-c/\tau+\frac{\sqrt{2}}{6}\{2\omega_3\omega_2^{-3/2}\right.\\&\qquad\left.+3\omega_2^{1/2}\alpha f''(\alpha)f'(\alpha)^{-1}\}x^2\right]\phi(x)+O(\theta_1^{-1})\end{aligned}$$

で与えられる．ここに $\tau=\{2\omega_2\alpha^2f'(\alpha)^2\}^{1/2}$ とする．この場合 $1/\sqrt{\theta_1}$ のオーダーで展開していることに注意する．従って x^2 の係数を 0 とおいた微分方程式

$$2\omega_3\omega_2^{-3/2}+3\omega_2^{1/2}\alpha f''(\alpha)f'(\alpha)^{-1}=0$$

を解くことによって

$$\left(\frac{Q_p}{\theta_1}\right)^{1-2\omega_3/(3\omega_2^2)}=\left(\frac{Q_p}{\theta_1}\right)^{1-2\theta_1\theta_3/(3\theta_2^2)}$$

なる変換形をえる．そこでこの変換形に対して $c=-2\omega_3f'(1)/(3\omega_2)$ で与えられる修正項を含めることによって

$$\Pr\left[\sqrt{\theta_1}\left\{\left(\frac{Q_p}{\theta_1}\right)^h-1-\frac{1}{\theta_1}h(h-1)(\theta_2/\theta_1)\right\}/\{\sqrt{2\theta_2/\theta_1}\,h\}<x\right]=\Phi(x)+O(\theta_1^{-1}) \tag{3.3}$$

を得る．ここに $h=1-2\theta_1\theta_3/(3\theta_2^2)$ とする．これが二次形式統計量に対する正規化変換である．

これに対して分散安定化変換は漸近分散 $\tau=(\sqrt{2\omega_2}\alpha f'(\alpha)=1$ の解すなわち

$$(2\theta_2/\theta_1)^{-1/2}\log(Q_p/\theta_1)$$

であり，この場合正規化変換は同時に分散安定化変換とはなりえないことがわかる．

次に (3.1) において $a_j=0$, $\lambda_j=1$；$j=1, 2, \cdots, p$ とおくと，$Q_p(x)$ は自由度 p の χ^2-分布となる．このとき (3.2) で定義される θ_r は，すべての r に対して $\theta_r=p$ となり，したがって (3.3) は

$$\Pr\left[\sqrt{\frac{9p}{2}}\left\{\left(\frac{Q_p}{p}\right)^{1/3}-1+\frac{2}{9p}\right\}<x\right]=\Phi(x)+O(p^{-1}) \tag{3.4}$$

の形に帰着される．すなわちこれは良く知られた χ^2-変量に対する Wilson and Hiferty (1931) の変換を示していることがわかる．

Wilson and Hiferty (1931) は変換形をベキ変換 $(Q_p/p)^h$ に限定し，$(Q_p/p)^h$ の平均のまわりの三次のモーメントの展開式の第一項が 0 となるように h を定めている．そこでは，変換形をベキ変換に限定しており，また (3.4) 式に含まれる $2/(9p)$ の項の意味づけがいま一つ明確でなかった．本稿では，変換の関数形は限定せず，極限分布への収束速度の改善という見地から

正規化変換を把えたが，これによって周知の二つの変換すなわち標本相関係数に対する Fisher の z 変換および χ^2-変量に対する Wilson-Hilferty の変換が，同じ規準のもとで導出されることがわかる．

尚，Jensen and Solomon (1972) は，Wilson-Hilferty の方法を用いて二次形式統計量の正規化変換 (3.3) を得ている．

3.2. 標本分散共分散行列の固有根

（i） 固有根の正規化変換

平均ベクトル μ, 分散共分散行列 Σ をもつ p 変量正規母集団からの大きさ $N=n+1$ 個の無作為標本にもとづく標本分散共分散行列を S とする．S の p 個の固有根を $l_1>l_2>\cdots>l_p>0$, Σ の固有根を $\lambda_1\geqq\lambda_2\geqq\cdots\geqq\lambda_p$ とする．このとき第 α 番目の固有根 l_α に対する正規化変換を求める．

一般式 (2.11) に相当する固有根 l_α の実数値関数 $f(l_\alpha)$ の分布の漸近展開は，もし対応する母集団固有根 λ_α が単根であれば

$$\Pr\{\sqrt{n}\{f(l_\alpha)-f(\lambda_\alpha)-c/n\}/\tau_\alpha<x]=\Phi(x)$$
$$-\frac{1}{\sqrt{2n}}\left[\sum_{\beta\neq\alpha}^{p}\frac{\lambda_\beta}{\lambda_\alpha-\lambda_\beta}-\frac{2}{3}-\frac{c\sqrt{2}}{\tau_\alpha}+\left\{\frac{2}{3}+\lambda_\alpha f''(\lambda_\alpha)f'(\lambda_\alpha)^{-1}\right\}x^2\right]\phi(x)+O(n^{-1})$$

で与えられる (Konishi (1977), Theorem 2.1)．ここに $\tau_\alpha=\sqrt{2}\lambda_\alpha f'(\lambda_\alpha)$ とする．一般論にしたがうと微分方程式

$$\frac{2}{3}+\lambda_\alpha f''(\lambda_\alpha)f'(\lambda_\alpha)^{-1}=0 \tag{3.5}$$

の解を求めればよいことになる．この解は，$l_\alpha^{1/3}$ で χ^2-変量に対する Wilson-Hilferty の変換と同じ立方根変換が導出されたことになる．そこでこの変換形に対する修正項 c を代入すると，固有根 l_α に対する正規化変換

$$\Pr\left[\sqrt{\frac{9n}{2}}\left\{\left(\frac{l_\alpha}{\lambda_\alpha}\right)^{1/3}-1-\frac{1}{3n}\left(\sum_{\beta\neq\alpha}^{p}\frac{\lambda_\beta}{\lambda_\alpha-\lambda_\beta}-\frac{2}{3}\right)\right\}<x\right]=\Phi(x)+O(n^{-1})$$

が求まる．一方分散安定化変換は漸近分散 τ に対する微分方程式

$$\tau_\alpha=\sqrt{2}\lambda_\alpha f'(\lambda_\alpha)=1$$

の解で，$(1/\sqrt{2})\log l_\alpha$ となる．したがって正規化変換とは関数形が異なる．これは (3.5) より

$$\frac{2}{3}+\lambda_\alpha f''(\lambda_\alpha)f'(\lambda_\alpha)^{-1}=\frac{2}{3}+\lambda_\alpha\{\log f'(\lambda_\alpha)\}'$$
$$=\frac{2}{3}+\lambda_\alpha\left\{\log\frac{1}{\sqrt{2}\lambda_\alpha}\right\}'\neq0$$

であることによる．

（ii） 固有根の比の正規化変換

標本分散共分散行列 S の第 i 番目および第 j 番目の固有根 l_i, l_j に対応する母集団固有根 λ_i, λ_j はどちらも単根であると仮定する．簡単のため $l_{ij}=l_i/l_j$, $\lambda_{ij}=\lambda_i/\lambda_j$ と略記する．このとき，l_{ij} の実数値関数 $f(l_{ij})$ の分布の漸近展開は，

$$\Pr\left[\sqrt{n}\left\{f(l_{ij})-f(\lambda_{ij})-\frac{1}{n}c\right\}/\tau_{ij}<x\right]$$
$$=\Phi(x)-\frac{1}{\sqrt{n}}\left[\frac{1}{2}\left(\sum_{\beta\neq i}^{p}\frac{\lambda_\beta}{\lambda_i-\lambda_\beta}-\sum_{\beta\neq j}^{p}\frac{\lambda_\beta}{\lambda_j-\lambda_\beta}\right)-c/\tau_{ij}\right.$$
$$\left.+\{1+\lambda_{ij}f''(\lambda_{ij})f'(\lambda_{ij})^{-1}\}x^2\right]\phi(x)+O(n^{-1})$$

で与えられる．ここに $\tau_{ij}=2\lambda_{ij}f'(\lambda_{ij})$ とする．したがって微分方程式

$$1+\lambda_{ij}f''(\lambda_{ij})f'(\lambda_{ij})^{-1}=0 \tag{3.6}$$

を解くことによって

$$f(l_{ij})=\frac{1}{2}\log l_{ij}$$

が求まる．この場合漸近分散 τ_{ij} に対して $\tau_{ij}=1$ を (3.6) に代入すると

$$1+\lambda_{ij}[\log\{f'(\lambda_{ij})\}]'=1+\lambda_{ij}\left[\log\frac{1}{2\lambda_{ij}}\right]'=0$$

であるから同様に分散安定化変換でもある．

一般論にしたがうと結局

$$\Pr\left[\frac{\sqrt{n}}{2}\left\{\log l_{ij}-\log\lambda_{ij}-\frac{1}{n}\left(\sum_{\beta\neq i}^{p}\frac{\lambda_\beta}{\lambda_i-\lambda_\beta}-\sum_{\beta\neq j}^{p}\frac{\lambda_\beta}{\lambda_j-\lambda_\beta}\right)\right\}<x\right]=\Phi(x)+O(n^{-1})$$

が求まる．

3.3　正準相関係数

$(p+q)$ 変量確率ベクトル x を二つの部分変量 $x=(x_1', x_2')'$ に分割する．ただし x_1, x_2 は各々 p 変量および q 変量確率ベクトルとする．いま $(p+q)$ 変量正規母集団からの大きさ $N=n+1$ 個の無作為標本にもとづく，x_1 と x_2 の間の標本正準相関係数を $1>r_1>r_2\cdots>r_p>0$，対応する母集団正準相関係数を $1>\rho_1\geqq\rho_2\geqq\cdots\geqq\rho_p>0$ とする．このときもし ρ_α が単根であれば，$r_\alpha{}^2$ の実数値関数 $f(r_\alpha^2)$ の分布の漸近展開は

$$\begin{aligned}&\Pr[\sqrt{n}\{f(r_\alpha^2)-f(\rho_\alpha^2)-c/n\}\tau_\alpha<x]\\&=\Phi(x)-\frac{1}{2\rho_\alpha\sqrt{n}}[p+q-2+\rho_\alpha^2+2(1-\rho_\alpha^2)\sum_{\beta\neq\alpha}^{p}\rho_\beta^2/(\rho_\alpha^2-\rho_\beta^2)-\frac{c}{\tau_\alpha}\\&\quad+\{1-3\rho_\alpha^2+2\rho_\alpha^2(1-\rho_\alpha^2)f''(\rho_\alpha^2)f'(\rho_\alpha^2)^{-1}\}x^2]\phi(x)+O(n^{-1}),\end{aligned}$$

で与えられる（Fujikoshi (1978)）．ここに $\tau=2\rho_\alpha(1-\rho_\alpha^2)f'(\rho_\alpha^2)$ とする．そこで微分方程式

$$1-3\rho_\alpha^2+2\rho_\alpha^2(1-\rho_\alpha^2)f''(\rho_\alpha^2)f'(\rho_\alpha^2)^{-1}=0. \tag{3.7}$$

の解を求めると $f(r^2)=(1/2)\log\{(1+r_\alpha)/(1-r_\alpha)\}$ を得る．それゆえ同様の議論によって

$$\Pr\left[\sqrt{n}\left(\frac{1}{2}\log\frac{1+r_\alpha}{1-r_\alpha}-\frac{1}{2}\log\frac{1+\rho_\alpha}{1-\rho_\alpha}-\frac{c_\alpha}{n}\right)<x\right]=\Phi(x)+O(n^{-1}) \tag{3.8}$$

が導出される．ここに $c_\alpha=\{p+q-2+\rho_\alpha^2+2(1-\rho_\alpha^2)\sum_{\beta\neq\alpha}^{p}\rho_\beta^2/(\rho_\alpha^2-\rho_\beta^2)\}/2\rho_\alpha$ とする．またこの場合 (3.7)，および漸近分散が $\tau_\alpha=2\rho_\alpha(1-\rho_\alpha^2)f'(\rho_\alpha^2)$ となることから

$$1-3\rho_\alpha^2+2\rho_\alpha^2(1-\rho_\alpha^2)\left[\log\frac{1}{2\rho_\alpha(1-\rho_\alpha^2)}\right]'=0$$

であり，したがって分散安定化変換も同じ関数によって達成されることがわかる．

3.4.　重相関係数，偏相関係数

（i）　重相関係数

$(q+1)$ 変量確率ベクトル $(x_1, x_2')'$ に対して，x_1, x_2 間で定義される母集団重相関係数 ρ_R をもつ $(q+1)$ 変量正規母集団からの大きさ $N=n+1$ 個の無作為標本にもとづく標本重相関係数を R とする．このとき (3.8) の特別な場合として

$$\Pr\left[\sqrt{n}\left\{\frac{1}{2}\log\frac{1+R}{1-R}-\frac{1}{2}\log\frac{1+\rho_R}{1-\rho_R}-\frac{1}{2\rho_R n}(q-1+\rho_R^2)\right\}<x\right]=\Phi(x)+O(n^{-1})$$

をえる．

(ii) 偏相関係数

$(p+q)$ 変量確率ベクトル $(x_1, \cdots, x_p, x_{p+1}, \cdots, x_{p+q})$ に対して，$x_i, x_j (1\leqq i, j\leqq p)$ 間の母集団偏相関係数を $\bar{\rho}_{ij}=\rho_{ij\cdot p+1,\cdots,p+q}$ 対応する標本偏相関係数を $\bar{r}_{ij}=r_{ij\cdot p+1,\cdots,p+q}$ とする．r_{ij} の分布は，二変量正規母集団からの大きさ $N-q$ 個の無作為標本にもとづく標本相関係数の分布と同じであるから，標本相関係数に関する議論から直ちに

$$\Pr\left(\sqrt{n-q}\left[\frac{1}{2}\log\{(1+\bar{r}_{ij})/(1-\bar{r}_{ij})\}-\frac{1}{2}\log\{(1+\bar{\rho}_{ij})/(1-\bar{\rho}_{ij})\}\right.\right.$$
$$\left.\left.-\bar{\rho}_{ij}/2(n-q)\right]<x\right)=\Phi(x)+O(n^{-1})$$

を得る．

補　　　　遺

2.2 節で用いた漸近正規性を有する統計量 T_n の実数値関数に対する漸近展開の導出法の概略を与える．2.2 節ではオーダー $1/\sqrt{n}$ の項までの漸近展開しか必要としなかったがここでは $\frac{1}{n}$ の項までの漸近展開を導出する．そのため実数値関数 $f(T_n)$ は，$T_n=\mu(\theta)(=\mu)$ の近傍で1対1かつ三回連続微分可能と仮定する．$f(T_n)$ を $T_n=\mu$ のまわりで Taylor 展開すると

$$f(T_n)=f(\mu)+(T_n-\mu)f'(\mu)+\frac{1}{2}(T_n-\mu)^2f''(\mu)+\frac{1}{6}(T_n-\mu)^3f^{(3)}(\mu)+R_n$$

ここに R_n は剰余項とする．

いま $\sqrt{n}\{f(T_n)-f(\mu)\}/\{\sigma(\theta)f'(\mu)\}$ の j 次キュムラントを κ_j とする．$f(T_n)$ の Taylor 展開式からわかるようにキュムラント κ_j を計算するには，$T_n-\mu$ の必要な次数のモーメントあるいはモーメントの展開式を求める必要がある．統計量 T_n は母集団から抽出された無作為標本 $x_1, x_2, \cdots, x_n$ の関数 $T_n=T_n(x_1, x_2, \cdots, x_n)$ であり，その期待値は漸近的に μ であることを仮定している．そこで T_n を μ のまわりで Taylor 展開し，項別に期待値をとることによってモーメントの展開式をもとめることができる．これがいわゆる "Delta Method" と呼ばれているもので種々の統計量のモーメントあるいはキュムラントの展開式を求めるのに有効な方法として用いられてきた．

このようにして導出された $\sqrt{n}\{f(T_n)-f(\mu)\}/\{\sigma(\theta)f'(\mu)\}$ のキュムラント κ_j は

$$\kappa_j=O(n^{-(j-2)/2})\qquad j\geqq 3$$

をみたすものと仮定し，

$$\kappa_1=\frac{1}{\sqrt{n}}\kappa_1^{(1)}+O(n^{-3/2}),\quad \kappa_2=1+\frac{1}{n}\kappa_2^{(2)}+O(n^{-2})$$
$$\kappa_3=\frac{1}{\sqrt{n}}\kappa_1^{(3)}+O(n^{-3/2}),\quad \kappa_4=\frac{1}{n}\kappa_2^{(4)}+O(n^{-2})$$

とおく．これらを $\sqrt{n}\{f(T_n)-f(\mu)\}/\{\sigma(\theta)f'(\mu)\}$ の特性関数

$$\psi(t)=\exp\left\{\sum_{j=1}^{\infty}(it)^j\kappa_j/j!\right\}$$

に代入し整理すると

$$\psi(t)=\exp\left(-\frac{1}{2}t^2\right)\cdot\exp\left[\frac{1}{\sqrt{n}}\left\{(it)\kappa_1^{(1)}+\frac{1}{6}(it)^3\kappa_1^{(3)}\right\}\right.$$
$$\left.+\frac{1}{n}\left\{\frac{1}{2}(it)^2\kappa_2^{(2)}+\frac{1}{24}(it)^4\kappa_2^{(4)}\right\}\right]+O(n^{-3/2})$$
$$=\exp\left(-\frac{1}{2}t^2\right)\left(1+\frac{1}{\sqrt{n}}\left\{(it)\kappa_1^{(1)}+\frac{1}{6}(it)^3\kappa_1^{(3)}\right\}\right.$$
$$+\frac{1}{n}\left[(it)^2\left\{\frac{1}{2}\kappa_2^{(2)}+(\kappa_1^{(1)})^2\right\}+(it)^4\left\{\frac{1}{24}\kappa_2^{(4)}+\frac{1}{3}\kappa_1^{(1)}\kappa_1^{(3)}\right\}\right.$$
$$\left.\left.+(it)^6\frac{1}{36}(\kappa_1^{(3)})^2\right]\right)+O(n^{-3/2})$$

となる．いま $\phi\{x\}$ を標準正規密度関数とし

$$\frac{1}{2\pi}\int(it)^j e^{-t^2/2}e^{-itx}dt=H_j(x)\phi(x)$$

を用いて上式を反転する．ただし $H_j(x)$ は

$$(-d/dx)^j\phi(x)=H_j(x)\phi(x)$$

であたえられるエルミート多項式とする．その結果，次の式を得る．

$$\Pr[\sqrt{n}\{f(T_n)-f(\mu)\}/\{\sigma(\theta)f'(\mu)\}<x]=\Phi(x)-\frac{1}{\sqrt{n}}\left\{\kappa_1^{(1)}\right.$$
$$\left.+\frac{1}{6}\kappa_1^{(3)}H_2(x)\right\}\phi(x)-\frac{1}{n}\left[\frac{1}{2}\{\kappa_2^{(2)}+(\kappa_1^{(1)})^2\}H_1(x)+\frac{1}{6}\left\{\frac{1}{4}\kappa_2^{(4)}\right.\right.$$
$$\left.\left.+\kappa_1^{(1)}\kappa_1^{(3)}\right\}H_3(x)+\frac{1}{72}(\kappa_1^{(3)})^2H_5(x)\right]\phi(x)+O(n^{-3/2})$$

したがって

$$a_1(\theta,f'(\mu),f''(\mu))=\kappa_1^{(1)}-\frac{1}{6}\kappa_1^{(3)},\quad a_2(\theta,f'(\mu),f''(\mu))=\frac{1}{6}\kappa_1^{(3)}$$

とおくと (2.11) の形の漸近展開式を得る．尚，(2.11) 式には関数 $f(T_n)$ の漸近平均に対する補正項 c が含まれているが，これは

$$\sqrt{n}\left\{f(T_n)-f(\mu)-\frac{1}{n}c\right\}/\{\sigma(\theta)f'(\mu)\}$$

の一次のキュムラントが

$$\kappa_1^*=\frac{1}{\sqrt{n}}[\kappa_1^{(1)}-c/\{\sigma(\theta)f'(\mu)\}]+O(n^{-3/2})$$

となることから，特性関数 $\psi(t)$ に含まれる $\kappa_1^{(1)}$ をあらためて $\kappa_1^{(1)}-c/\{\sigma(\theta)f'(\mu)\}$ と置き換えることによって求めたものである．

以上のような漸近展開は，適当な正則条件が満たされれば有効であるということが Wallace (1958) によって予想され，その後 Bhattacharya and Ghosh (1978) によって，より精密に定式化された形で証明された．本稿で議論した多変量解析における各種統計量の分布の漸近展開の有効性は主として Bhattacharya and Ghosh (1978) の Theorem 2 に基づくものである．

謝　辞

この研究をまとめるにあたり，有益なコメントをいただいた仁木直人氏に，また的確な助言をいただいた査読者の皆様に感謝いたします．

参 考 文 献

Anderson, T.W. (1958). *An Introduction to Multivariate Analysis,* Wiley, New York.

Bartlett, M.S. (1948). The use of transformations, *Biometrics,* **3**, 39-52.

Bhattacharya, R.N. and Ghosh, J.K. (1978). On the validity of the formal Edgeworth expansion, *Ann. Statist.,* **6**, 434-451.

Efron, B. (1982). Transformation theory : How normal is a family of distributions ?, *Ann. Statist.,* **10**, 323-339.

Fisher, R.A. (1921). On the 'probable error' of a coefficient of correlation deduced from a small sample, *Metron,* **1**, 1-32.

Fisher, R.A. (1958). *Statistical Methods for Research Workers,* Oliver and Boyd, Edinburgh.

Fujikoshi, Y. (1978). Asymptotic expansions for the distributions of some functions of the latent roots of matrices in three situations, *J. Multivariate Anal.,* 8, 63-72.

Hotelling, H. (1953). New light on the correlation coefficient and its transforms, *J.R. Statist. Soc.,* B, **15**, 193-232.

Jensen, D.R. and Solomon, H. (1972). A Gaussian approximation to the distribution of a definite guadratic form. *J. Amer. Statist, Ass.,* **67**, 898-902.

Johnson, N.L. and Kotz, S. (1970). *Distributions in Statistics : Continuous Univariate Distributions*-2, Boston : Houghton Mifflin.

Konishi, S. (1977). Asymptotic expansion for the distribution of a function of latent roots of the covariance matrix, *Ann. Inst. Statist. Math.,* **29**, 389-396.

Konishi, S. (1978). An approximation to the distribution of the sample correlation coefficient, *Biometrika,* **65**, 654-656.

Konishi, S. (1981). Normalizing transformations of some statistics in multivariate analysis, *Biometrika,* **68**, 647-651.

Konishi, S. (1984). Normalizing and variance stabilizing transformations for intraclass correlations, To appear in *Ann. Inst. Statist. Math.*

Kraemer, H.C. (1973). Improved approximation to the non-null distribution of the correlation coefficient, *J. Amer. Statist. Ass.,* **68**, 1004-1008.

Niki, N. and Konishi, S. (1984 a). Higher order asymptotic expansions for the distribution of the sample correlation coefficient, *Commun. Statist. Simula. Computa.,* B13, **3.**

Niki, N. and Konishi, S. (1984 b). Normalizing transformations for higher order asymptotic expansions, *Research Memo.,* No. **284**, The Institute of Statistical Mathematics.

Ruben, H. (1966). Some new results on the distribution of the sample correlation coefficient, *J.R. Statist. Soc.,* B, **28**, 513-525.

Wallace, D.L. (1958). Asymptotic approximations to distributions, *Ann. Math. Statist.,* **29**, 635-654.

Wilson, E.B. and Hilferty, M.M. (1931). The distribution of chi-square, *Proc. Nat. Acad. Sci.,* **17**, 684-688.

Normalizing and Variance Stabilizing Transformations of Multivariate Statistics

Sadanori Konishi

(The Institute of Statistical Mathematics)

A classical and practical example of transformations of statistics is Fisher's z transformation for the correlation coefficient in a bivariate normal sample. Transformations of this kind are made for two different reasons; one is to normalize an asymptotic distribution, and another is to stabilize an asymptotic variance. It is well known that Fisher's z transformation possesses these two properties.

There is a question unanswered: why the normalization and variance stabilization can be simultaneously achieved by the same transformation. One object of this paper is to answer the question theoretically.

In multivariate analysis there are many statistics whose limiting distributions are normal. So, it is of interest to make a suitable transformation in order to obtain simple and accurate approximations to the distributions of statistics. After constructing a concept of normalization, a general procedure for finding a normalizing transformation is given and applied to various statistics in multivariate analysis. The relationship between normalization and variance stabilization is also discussed.

Key words: Asympototic expansions, Fisher's z transformation, multivariate statistics, normalizing transformation, variance stabilizing transformation.